APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Nov. 2005, p. 7253–7262 Vol. 71, No. 11
0099-2240/05/$08.00+0 doi:10.1128/AEM.71.11.7253–7262.2005

In Silico Reconstruction of the Metabolic Pathways of *Lactobacillus plantarum*: Comparing Predictions of Nutrient Requirements with Those from Growth Experiments

Bas Teusink,[1,2,3]† Frank H. J. van Enckevort,[2,3]† Christof Francke,[1,3]† Anne Wiersma,[1,2] Arno Wegkamp,[1,2] Eddy J. Smid,[1,2] and Roland J. Siezen[1,2,3]*

Wageningen Centre for Food Sciences, P.O. Box 557, 6700AN Wageningen, The Netherlands[1]; NIZO food research, P.O. Box 20, 6710BA Ede, The Netherlands[2]; and Centre for Molecular and Biomolecular Informatics, Radboud University, P.O. Box 9010, 6500GL Nijmegen, The Netherlands[3]

Received 12 April 2005/Accepted 20 June 2005

On the basis of the annotated genome we reconstructed the metabolic pathways of the lactic acid bacterium *Lactobacillus plantarum* WCFS1. After automatic reconstruction by the Pathologic tool of Pathway Tools (http://bioinformatics.ai.sri.com/ptools/), the resulting pathway-genome database, LacplantCyc, was manually curated extensively. The current database contains refinements to existing routes and new gram-positive bacterium-specific reactions that were not present in the MetaCyc database. These reactions include, for example, reactions related to cell wall biosynthesis, molybdopterin biosynthesis, and transport. At present, LacplantCyc includes 129 pathways and 704 predicted reactions involving some 670 chemical species and 710 enzymes. We tested vitamin and amino acid requirements of *L. plantarum* experimentally and compared the results with the pathways present in LacplantCyc. In the majority of cases (32 of 37 cases) the experimental results agreed with the final reconstruction. LacplantCyc is the most extensively curated pathway-genome database for gram-positive bacteria and is open to the microbiology community via the World Wide Web (www.lacplantcyc.nl). It can be used as a reference pathway-genome database for gram-positive microbes in general and lactic acid bacteria in particular.

Lactic acid bacteria (LAB) are members of the low-G+C-content gram-positive bacteria and play a pivotal role in modern biotechnology. They are used in starter cultures for food fermentations, where their metabolic activities contribute to food preservation, taste, and texture development. Increasingly LAB are also used, through metabolic engineering, for the production of flavor compounds (17, 28), vitamins, such as riboflavin (7) and folic acid (10, 18, 51), or even more complex molecules, such as exopolysaccharides (4). Some strains are marketed as probiotics, which are claimed to positively affect human and/or animal health. Moreover, the persistence of some LAB species in the gastrointestinal tract, their amenability to genetic manipulation, and their proven safety through use in food production processes for thousands of years offer possibilities for vaccine production and delivery of therapeutic agents (48).

In the genomics era, attention is shifting from empirical strain selection and enzyme characterization to genome sequencing and functional genomics (1, 24, 27, 41, 47). We recently sequenced the 3.3-Mbp genome of *Lactobacillus plantarum* WCFS1 (26), a versatile and flexible lactic acid bacterium that is used in food fermentations and is sold as a probiotic (30). The complete genome sequence was analyzed, and putative biological functions were assigned to 2,120 (70%) of the 3,052 predicted protein-encoding genes.

* Corresponding author. Mailing address: Centre for Molecular and Biomolecular Informatics, Radboud University, P.O. Box 9010, 6500GL Nijmegen, The Netherlands. Phone: 31 243653379. Fax: 31 24 365 29 77. E-mail: Roland.Siezen@nizo.nl.

† B.T., F.H.J.V.E., and C.F. contributed equally to this work.

Annotation of the genome, however, is only the first step toward understanding the function of genes. To understand biological function, individual genes have to be put into a biological context, including the context of metabolic pathways. Reconstruction of metabolic pathways is therefore imperative for in silico predictions of gene function and the metabolic (potential) capabilities of an organism. Many databases and tools that aid metabolic reconstruction can be found on the World Wide Web (see the annual January issue of Nucleic Acids Research for an overview of databases in bioinformatics). These resources differ considerably in scope, visualization possibilities, and organism specificity (see reference 47 for tools relevant to LAB). Unfortunately, most of the metabolic databases either are not organism specific or have been generated automatically and hence contain many errors. There are, however, well-curated databases, and EcoCyc is the prime example (21). For lactic acid bacteria, and even for gram-positive bacteria in general, such a high-quality database has not been available. We therefore decided to construct a well-curated metabolic database for a gram-positive bacterium, *L. plantarum*.

In this paper we describe the reconstruction of the metabolic pathways of *L. plantarum* WCFS1 within the framework of Pathway Tools (19, 20). We constructed a pathway-genome database, which we called LacplantCyc and which is publicly available. The consistency of the reconstructed metabolic pathways for vitamin and amino acid biosynthesis was tested by systematic comparison with growth requirements. The analysis raised some interesting questions that are also described below.

MATERIALS AND METHODS

Generation of LacplantCyc. Pathway Tools and Pathologic, version 8.0, were installed on a Linux server (Dell Poweredge 2600, Redhat 8.0). The Pathologic software of Pathway Tools (19) and the MetaCyc database (29) were used to automatically generate a pathway-genome database (PGDB) from the complete genome of *L. plantarum* WCFS1 (GenBank file ID NC-004567). The Pathologic software automatically associates genes with reactions, based on enzyme code (EC) numbers and function descriptions contained in the GenBank file of the annotated genome. The PGDB comprises the complete genome content of *L. plantarum* WCFS1 and connections between the coding sequences, potential genes, enzymes, reactions, and biochemical pathways. For simplicity we most often refer to coding sequences as genes below.

Functional annotation. For proper functional annotation, the distinction between paralogs and orthologs is important. To distinguish between orthologs and paralogs, we used the following procedure. In each instance in which the original annotation of a gene was reconsidered (e.g., for most transporters), we performed a BLAST search with the corresponding protein sequence in the ERGO database (www.integratedgenomics.com) against all available bacterial genomes or an appropriate subset (e.g., all firmicute genomes). We collected the coding sequences with E values less than 1×10^{-5} and aligned them with CLUSTALX (54) (not more than 200 sequences) or Muscle (11). The alignment was inspected by eye, deviant sequences were removed, and a bootstrap neighbor-joining tree was generated. The tree was divided into clusters on the basis of branching. For each cluster, single species representatives were considered orthologous and therefore to have identical molecular functions. When the molecular function of one of the proteins in the cluster was experimentally verified, that function was transferred to all members of the group. Species representatives in neighboring clusters were considered paralogous, and molecular functions identical to that of the paralog were assumed, unless additional information was available that suggested otherwise. Genome context was used to verify the orthology/paralogy assignment and to provide additional information on molecular function.

Single-omission growth experiments. A chemically defined medium (CDM) for growth of *L. plantarum* was developed based on the study of Poolman and Konings (39). This medium contains vitamins, seven nucleotides, and 18 amino acids and is described in Table 1. *L. plantarum* WCFS1 grows well on CDM; the maximal growth rate is about 0.6 h^{-1}, and the final optical density at 600 nm is 5 (results not shown). In single-omission experiments, each of the vitamins and amino acids was separately left out of the medium, and growth was monitored in 10-ml tubes at 37°C under microaerobic conditions (i.e., without stirring).

RESULTS AND DISCUSSION

In silico reconstruction. Using the GenBank file of the annotated genome of *L. plantarum* WCFS1 as input, the Pathologic software automatically assigned 211 pathways and 915 reactions (Table 2). The log file generated indicated that 659 coding sequences were assigned a reaction on the basis of matching EC numbers. For 556 coding sequences both the EC number and the function name in the GenBank file perfectly matched the MetaCyc entry, whereas for 103 coding sequences they did not. The "mismatch" in almost all cases appeared to be a problem of synonyms lacking in the MetaCyc database or of nonspecificity of the EC number (mainly EC 2.7.1.69) (see below). In addition, 66 coding sequences without a complete EC number in the GenBank file were connected to a reaction based on matching the function name. A total of 725 enzyme-encoding genes were automatically matched to a reaction.

Subsequent curation of the automatic reconstruction involved (i) identifying and resolving inconsistencies (e.g., filling gaps in pathways), (ii) defining the often complex relationship between genes, proteins, and reactions, and (iii) introducing *Lactobacillus*-specific reactions and pathways.

(i) Identifying and resolving inconsistencies. The curation process involved comparing the automatic assignment of pathways by the Pathologic software with the pathways in databases such as KEGG (http://www.genome.jp/kegg/) and WIT (via the ERGO bioinformatics suite) and resolved possible discrepancies. We found that most pathways in our initial PGDB contained gaps (missing links); i.e., they contained reactions to which no coding sequences had been assigned. In many instances the source of a "missing link" was evident as the information necessary to attribute a particular reaction to a specific gene was absent from the GenBank file. In other cases it appeared that although the gene product had been properly annotated, it was not detected by the Pathologic software. These detection failures could often be attributed to the use of EC numbers in the automatic reconstruction procedure, as quite a few reactions lack a complete EC number (e.g., EC 3.6.1.-) (see below) and some EC numbers lack specificity (the most obvious example is EC 2.7.1.69, which refers to all protein phosphohistidine sugar phosphotransferase systems, representing about 60 genes in *L. plantarum*). Obviously, EC numbers were not designed for making automatic metabolic reconstructions. There are, however, other efforts in standardization, such as Gene Ontology (2), Systems Biology Markup Language (16), and Biopax (http://www.biopax.org/), that hopefully will provide a labeling system that is more suitable for automatic reconstruction, provided that a genome is annotated in these terms and that metabolic reaction databases, such as MetaCyc, recognize the labels.

Other, more complicated reasons for not being able to identify the genes belonging to certain steps in pathways include (i) the presence of analogous genes (i.e., nonhomologous genes with the same molecular function) or paralogous genes (i.e., groups of homologous genes encoding similar or identical molecular functions), (ii) reactions for which no gene has been identified yet in any organism, and (iii) genes encoding enzymes for which the substrate is difficult to predict based on homology. We are performing a systematic "reversed" functional annotation (i.e., looking for specific genes belonging to missing molecular functions) by looking for identified genes in related organisms and searching for orthologous genes in the *L. plantarum* genome. More recent versions of the Pathologic software have implemented an automatic procedure for identifying missing enzymes during the automatic reconstruction process (13). Finding missing enzymes is quite an exhaustive exercise and is beyond the scope of this paper (see reference 35 for a review). Here we give two examples. First, the fact that lysylated phosphatidylglycerol is a major component of the cell membrane of *L. plantarum* (34) led to a search and identification of lp_0690 as *mprF* (33), which encodes a lysylphosphatidyl-glycerol synthetase. Second, the molecular function of the lp_2263 gene product could be inferred from the conserved genome context and the need for an *N*-acetyldiaminopimelate deacetylase (EC 3.5.1.47) in the lysine biosynthesis route.

(ii) Defining the often complex relationships between genes, proteins, and reactions. Some enzymes can catalyze many reactions (e.g., alcohol dehydrogenase [EC 1.1.1.1]), whereas some single reactions are catalyzed by a large complex of proteins (e.g., ATP synthase). These many-to-many relationships between genes, proteins, and reactions are important for data analysis and visualization. Moreover, predicting the impact of deletions of metabolic genes, mapping of gene data to reaction fluxes, and visualization of data on metabolic maps require these relationships (8, 36). Pathway Tools allows specification of many-to-many relationships with Boolean logic, i.e.,

TABLE 1. Comparison of growth requirements and pathway information stored in LacplantCyc[a]

CDM component[b]	Concn (g/liter)	μ_{max}/OD_{600} at 24 h (% in CDM)	Initial LacplantCyc possible	Final LacplantCyc possible
Vitamins				
Ca-(D)-(+)-pantothenate (vitamin B_5)	0.001	19/3	−	−
D-Biotin (vitamin B_7)[c]	0.0025	85/21	−	−
Folic acid (vitamin B_{11})	0.001	100/99	−	?
Lipoic acid (6,8-thioctic acid)	0.001	100/105	−	−
Nicotinic acid[c]	0.001	5/1	−	−
D-Aminobenzoic acid	0.01	98/98	−	?
Pyridoxamine HCl	0.005	94/64	−	−
Pyridoxine HCl (vitamin B_6)	0.002	99/108	−	−
Pyridox-x[c]		81/34	−	−
Riboflavin (vitamin B_2)	0.001	79/14	−	−
Thiamine HCl (vitamin B_1)	0.001	99/91	−	−
Amino acids				
Alanine	0.24	97/95	−	+
Arginine	0.125	8/2	−	+
Aspartic acid	0.42	102/110	+	+
Cysteine-HCl	0.13	82/92	+	+
Glutamic acid	0.5	2/1	+	−
Glycine	0.175	44/68	+	+
Histidine	0.15	92/93	−	+
Isoleucine	0.21	61/23	−	−
Leucine	0.475	6/3	−	−
Lysine	0.44	98/105	−	+
Methionine	0.125	23/41	+	+
Phenylalanine	0.275	72/44	−	−
Proline	0.675	100/104	+	+
Serine	0.34	102/95	−	+
Threonine	0.225	97/98	+	+
Tryptophan	0.05	19/20	+	+
Tyrosine	0.25	85/70	−	+
Valine	0.325	6/4	−	−
Nucleotides[d]				
Adenine	0.01	+	+	+
Guanine	0.01	+	+	+
Inosine	0.005	+	+	+
Orotic acid (vitamin B_{13})	0.005	+	+	+
Thymidine	0.005	+	+	+
Uracil	0.01	+	+	+
Xanthine	0.01	+	+	+

[a] The first two columns indicate the compounds present in CDM and their concentrations. The compounds present in minimal medium are indicated by boldface type. The third column indicates the results of the single-omission growth experiments. Both the maximal growth rate (μ_{max}) and the optical density at 600 nm (OD_{600}) at 24 h after inoculation are shown, expressed as percentages of those in complete CDM. The last two columns indicate the potential ability of *L. plantarum* to synthesize the different compounds from central metabolic intermediates, as inferred from the presence (+) or absence (−) of the necessary enzymes in the automatically generated (column 4) and manually curated (column 5) versions of LacplantCyc. The question marks indicate uncertainties that are discussed in the text.

[b] The additional medium components were K_2HPO_4 (1 g/liter), KH_2PO_4 (5 g/liter), sodium acetate (1 g/liter), ammonium citrate (0.6 g/liter), glucose (10 g/liter), and ascorbic acid (vitamin C) (0.5 g/liter).

[c] These vitamins resulted in intermediate growth after transfer from CDM. To eliminate the possibility that intracellular dilution of a vitamin caused residual growth, cells were reinoculated into the same medium. Data from the second growth curve are shown. Pyridox-x indicates that both pyridoxamine and pyridoxine were omitted in the experiment.

[d] All nucleotides were omitted from the medium at the same time.

AND and OR relationships, where the OR relationship symbolizes the presence of isoenzymes and the AND relationship symbolizes the participation of proteins in protein complexes.

However, no information on the composition of protein complexes was extracted automatically by the Pathologic software; none of the nine protein complexes initially assigned was connected to a coding sequence. We identified and added 115 new protein complexes involving about 210 coding sequences to LacplantCyc. In EcoCyc most complexes are homo-oligomeric. As data on such complexes are mostly lacking for *L. plantarum*, we introduced into LacplantCyc only protein complexes consisting of two or more gene products. Since homo-oligomeric protein complexes are linked to a single gene, no essential information is left out when they are omitted. Reactions facilitated by protein complexes quite often can be subdivided into partial reactions catalyzed by individual subunits, and sometimes even the partial reactions have a separate EC number. We chose to leave out the partial reactions in cases where the intermediates were not used in other pathways. An example is the production of malonyl coenzyme A (malonyl-CoA) from acetyl-CoA in the initiation of fatty acid biosynthesis (in the pathway "fatty acid biosynthesis, initiation"), in which we lumped three partial reactions.

Automatic information transfer from MetaCyc to Lacplant-

TABLE 2. Comparison of the contents of LacplantCyc before and after manual curation

Parameter	After automatic reconstruction			After manual curation		
	No.	Connected to a gene	Occupancy (%)	No.	Connected to a gene	Occupancy (%)
Pathways[a]	211	≤1/3, 21%; 1/3–2/3, 34%; ≥2/3, 45%		129	≤1/3, 0%; 1/3–2/3, 5%; ≥2/3, 95%	
Reactions in pathways[b]	1,480	847	57	703	656	93
Unique reactions in PGDB[c]	915	476 (in, 338; out, 138)	48	704	652 (in, 621; out, 31)	93
Complexes[d]	9	0	0	119	51 (metabolism), 64 (transport)	97
Genes involved[e]	0			210		

[a] For each pathway in the PGDB the ratio of the number of gene-associated reactions to the total number of reactions in the pathway was determined. The third and sixth columns indicate the percentages of pathways with ratios less than or equal to 1/3, between 1/3 and 2/3, and equal to or more than 2/3.
[b] For each pathway the number of reactions and the number of gene-associated reactions were counted, and the percentage of the two was calculated (occupancy). A relatively high number of reactions recovered from the pathways compared with the total number of reactions in the database reflects a redundancy of reactions in the pathways.
[c] All reactions in the PGDB were inspected for occurrence within a pathway (columns 2 and 5) and for association with a gene (columns 3 and 6), where in indicates the number of reactions present within pathways and out indicates the number of reactions that were associated with a gene but not with a pathway.
[d] Number of complexes in the database.
[e] Number of genes associated with complexes.

Cyc was ineffective in the case of transport reactions because Pathway Tools was not designed to do this. Proper annotation of transport proteins is hampered by the fact that bacterial genomes include considerable numbers of paralogous genes related to transport. In fact, the two largest transporter classes in *L. plantarum* WCFS1 comprise at least 235 coding sequences. These systems, the ABC transporters and the phosphotransferase systems (PTSs), employ a sequential mechanism involving several protein-protein interactions to free the energy of a phosphoryl bond (from ATP and phosphoenolpyruvate, respectively), enabling the transport of certain nutrients (carbohydrates in the case of PTSs) or other molecules over the cell membrane (9, 40). Recently, we devised a strategy to aid the classification and annotation of paralogous systems, which is based primarily on the combination of phylogeny and gene context conservation (see Materials and Methods) (Francke et al., unpublished data). This strategy was applied to the active transport systems of *L. plantarum* WCFS1. The membrane transport mechanisms, like those mediated by ABC transporters and PTSs, were introduced into LacplantCyc as single reactions catalyzed by a protein complex. At this time LacplantCyc includes 11 PTS-mediated transport reactions (50 genes) and 31 ABC transporter-mediated reactions (80 genes).

Phosphotransfer mechanisms that are not related or only indirectly related to transport were dealt with in a similar fashion. An example is the phosphorylation of dihydroxyacetone in the cytosol. In *Escherichia coli*, this reaction is mediated by the general PTS proteins EI and HPr and a protein complex composed of DhaK, DhaL, and DhaM (14). Orthologous genes were identified in the *L. plantarum* genome, and the phosphotransfer mechanism was incorporated as a single reaction into LacplantCyc (Fig. 1).

(iii) Defining *Lactobacillus*-specific reactions and pathways. To make adjustments to reactions or pathways or to add new pathways, we first defined some rules for reasons of consistency. (i) A MetaCyc pathway was replaced by an *L. plantarum*-specific pathway whenever we made adjustments to the MetaCyc pathway (e.g., after a reaction was added or removed), and the names for the new pathways were chosen so that they were straightforward (e.g., "purine interconversions: adenine"). (ii) General pathways were made specific, but at the same time the

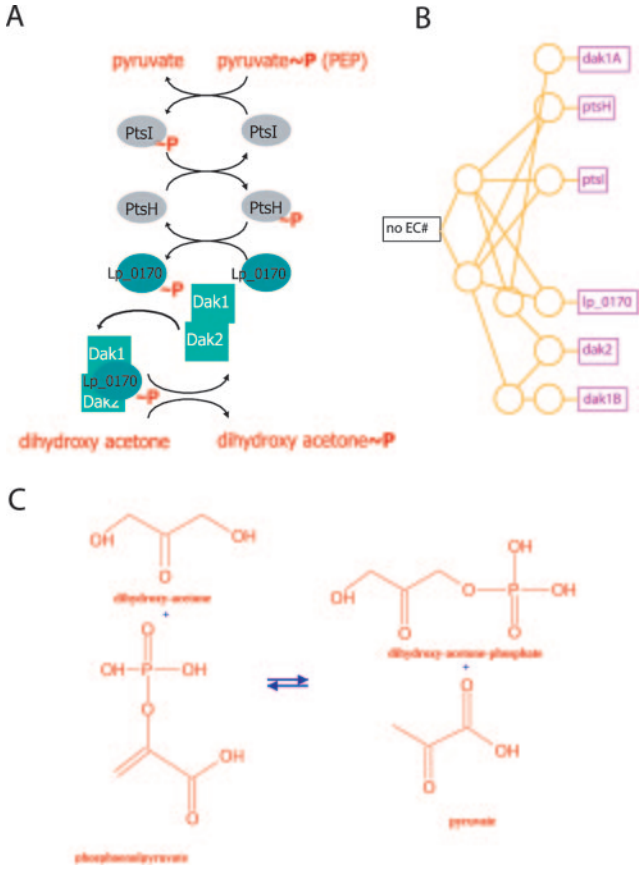

FIG. 1. Phosphorylation of dihydroxyacetone occurs via a phosphotransfer mechanism involving components of the PTS (see text for details). (A) Schematic representation of the biochemistry, with some proteins involved as phosphocarriers (ovals) and other proteins involved as catalysts (squares). (B) Gene-protein-reaction relationships for the system, as given by LacplantCyc. (C) Overall reaction as implemented in LacplantCyc. Purple boxes indicate genes that form a one-to-one relationship with an orange circle, the gene product (protein). The gene products can form protein complexes, which are indicated by convergence from two or more circles to a single circle. Finally, the protein or protein complex is associated with a reaction, indicated by a black box containing the EC number if it exists (in this example there is no EC number). Note that because of the presence of two *dak1* genes (*dak1A* and *dak1B*), two complexes carrying out the same reaction were defined.

generalized reactions were maintained (for instance, in lipid biosynthesis there are now two "fatty acid elongation" pathways, a general pathway with elongation expressed as $n \rightarrow n + 1$ and a complete pathway with each compound specified). (iii) In purine and pyrimidine synthesis ATP was used as a standard phosphoryl (energy) donor, and the possible use of, for example, GTP was mentioned in a comment. (iv) The occurrence of the same reaction in different pathways was restricted by adjusting pathway definitions (e.g., length). Without such restrictions, many pathways known to be absent in *L. plantarum*, such as the tricarboxylic acid (TCA) cycle, would be included, giving the false impression of many gaps in many pathways. (v) Similar and predominantly empty pathways, like the different versions of glycolysis in which the core is identical but the input and output reactions vary, were removed. (vi) Due to the inherent reversibility of any chemical reaction, enzymes that are normally assigned a function in biosynthesis were also considered to play a role in degradation. Consequently, we used the attribute "conversion" in the names of amino acid conversion pathways in LacplantCyc.

We made extensive rearrangements in the pathways for fatty acid biosynthesis and purine and pyrimidine biosynthesis. We also added a number of pathways that were absent from the MetaCyc database. These include citrate lyase formation, phosphoenolpyruvate:dihydroxyacetone phosphotransfer, and molybdopterin biosynthesis. Molybdopterin is a cofactor for a diverse group of enzymes involved in redox reactions, including nitrate reductase (55), an enzyme that is present in *L. plantarum* WCFS1. The four subunits of nitrate reductase are encoded by genes which are colocalized with the molybdopterin biosynthesis gene cluster (Fig. 2A). Although the compounds, enzyme names, and reactions for molybdenum cofactor biosynthesis are present in the MetaCyc database, they had not yet been assembled into a pathway. Figure 2B shows the pathway that we added to LacplantCyc, based on the primary literature (42). This pathway consists of four reactions. It was found that three of these reactions are catalyzed by a protein complex. It is also possible that all reactions take place in one large complex. GTP acts as a precursor, and two sulfur atoms are transferred to the molecule via a cysteine-derived thiol donor that shares chemistry with the thiol donor acting in thiamine biosynthesis (see below). Subsequently, the molybdenum metal ion is incorporated, and finally, the molybdenum cofactor matures.

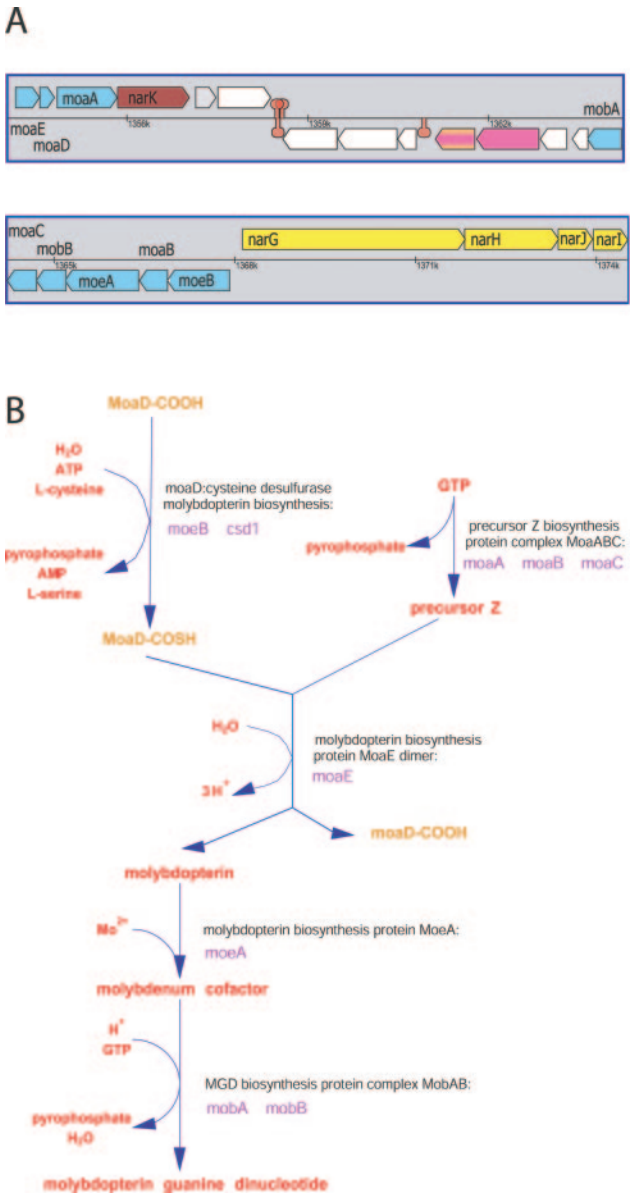

FIG. 2. Molybdenum cofactor biosynthesis: genes and pathway. (A) The molybdenum cofactor biosynthesis genes are located adjacent to nitrate reductase genes in the genome. (B) Molybdenum cofactor biosynthesis route as depicted in LacplantCyc.

Overview of the impact of manual curation. To evaluate the results of our curation procedure, we compared the automatically generated PGDB with the manually curated PGDB (Table 2). The numbers illustrate the redundancy in the automatically generated PGDB. Pathway Tools was purposefully designed to suggest many pathways so that no pathway is overlooked, but the downside of this design principle is extensive redundancy, as illustrated by the fact that of the 847 reactions that were found in all pathways and that were connected to a gene, only 338 were in fact unique (Table 2). Moreover, in 55% of the pathways less then two-thirds of the reactions were associated with a gene.

Table 2 shows that we succeeded in removing most of the redundancy, yet maintained the stored reaction-gene associations. Moreover, we added 210 new gene-protein complex, 176 new gene-reaction, and 283 reaction-pathway associations. The current version of LacplantCyc contains 129 pathways, compared to 211 pathways in the automatically generated PGDB. Initially, 57% of the pathway reactions were connected to a gene or multiple genes; in the current version this percentage is 93%. For comparison, the value for the current version of EcoCyc is also 93% (23), compared to 60% when an *E. coli* PDGD was automatically created with the Pathologic software based solely on the NCBI GenBank file (results not shown).

Predictions of biosynthetic capabilities made on the basis of LacplantCyc: strengthening pathway information with experimental data. Using the pathway information stored in LacplantCyc, we made predictions about the ability of *L. plantarum* WCFS1 to synthesize amino acids, nucleotides, and

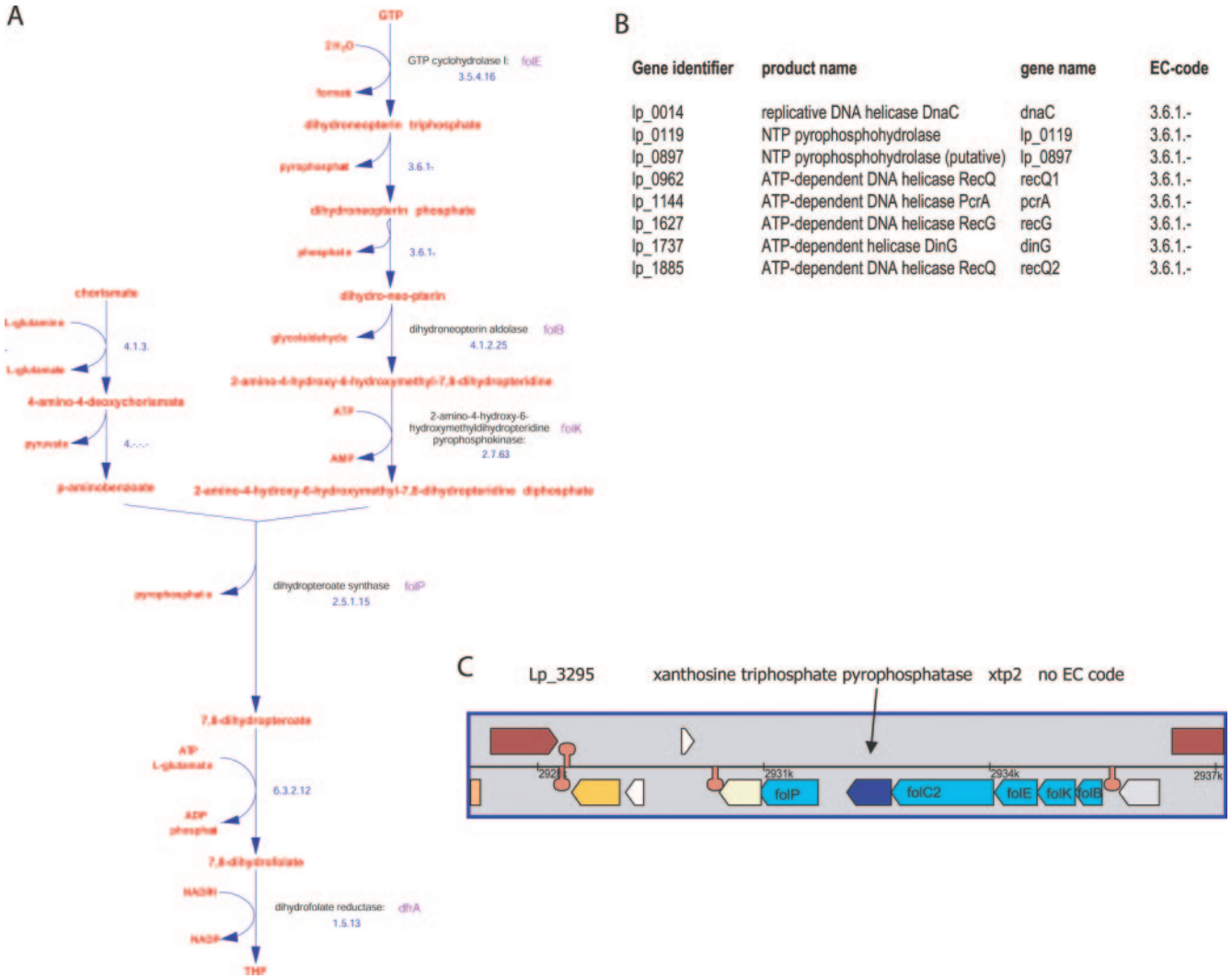

FIG. 3. Reconstruction of the folate biosynthesis pathway. (A) Original pathway reconstruction by Pathway Tools. (B) List of genes with the annotation EC 3.6.1.-, as described by Kleerebezem et al. (26). NTP, nucleoside triphosphate. (C) Gene context and annotation of the lp_3295 gene.

certain vitamins by looking at the presence or absence of crucial steps in the biosynthesis pathways. Simultaneously, single-omission growth experiments were performed (see Materials and Methods). The results of both analyses are summarized in Table 1. For only 5 of the 37 compounds evaluated were the predictions made on the basis of the reconstruction not in agreement with the experimental results. Nucleotide biosynthesis pathways were complete, and all nucleotides and precursors of nucleotides could indeed be omitted from the medium altogether (Table 1). When all nucleotides were omitted from the medium, a longer lag phase was observed, and the maximum growth rate decreased to 64% of that observed for complete medium, but the final optical density was similar. For vitamins and amino acids, further investigation was required for a few cases, as described below.

(i) Vitamins. Addition of the B vitamins folic acid (vitamin B_{11}), thiamine (vitamin B_1), and pyridoxine/pyridoxamine (vitamin B_6) to the medium was not required for growth, indicating that the pathways related to the biosynthesis of these compounds should be complete and active. Yet, initially these pathways appeared to be incomplete. We succeeded in filling most gaps in these pathways by combining sequence information from different species with previously published data.

In the first instance, Pathway Tools indicated that there were five gaps in the folate biosynthesis pathway (Fig. 3A), whereas the corresponding data for KEGG showed that there were only two gaps (related to *para*-aminobenzoate production [see below]). The first enzyme that was missing (the associated gene was not identified) was dihydroneopterin triphosphate pyrophospohydrolase (50). The corresponding reaction has an incomplete EC number (EC 3.6.1.-). Incomplete EC numbers are (justifiably) not accepted in the automatic reconstruction by the Pathologic software because of their ambiguity, but they are accepted by KEGG. Although in both the *L. plantarum* GenBank file and KEGG EC 3.6.1.- is associated with eight genes (Fig. 3B), none of the genes appeared to be appropriate for the reaction based on the annotated function names (a perfect illustration of the sometimes problematic nature of

automatic reconstruction). However, analysis of the genome content revealed that the *L. plantarum* genome does contain a suitable candidate for the molecular function. The corresponding gene (lp_3295) is located in the middle of a folate operon (Fig. 3C) and was annotated as the xanthosine pyrophosphatase gene (*xtp2*), yet it lacked an EC number. In *Lactococcus lactis*, a (nonhomologous) gene encoding a pyrophosphatase is also located in the folate gene cluster, and the gene product was recently shown to play the missing role in folate biosynthesis (25).

The subsequent enzyme in the pathway is dihydroneopterin phosphatase, and the reaction has the same EC number, EC 3.6.1.-. The reaction may be carried out by the same enzyme, but it could also be accommodated by nonspecific phosphomonoesterases (6), and there are many of these enzymes in the genome.

A third predicted gap in the folic acid biosynthesis pathway was the 7,8-dihydropteroate:L-glutamate ligase reaction (EC 3.6.2.12) catalyzed by FolC. This enzyme adds glutamate molecules to the pterin core, but depending on whether a glutamate is already present, the reaction is defined as EC 6.3.2.12 (glutamate ligase) or EC 6.3.2.17 (polyglutamate ligase). The corresponding gene was not automatically recognized by Pathway Tools, because the GenBank file contained only the latter EC number and this number was absent from the MetaCyc database.

The last two gaps in the folate biosynthesis pathway are associated with the conversion of chorismate to the *para*-aminobenzoate moiety that is attached to the pterin core to form 7,8-dihydropteroate. In *E. coli* the conversion is a two-step process catalyzed by the PabA-PabB complex and PabC. However, orthologs of these proteins are absent from *L. plantarum* and all other LAB sequenced (with the exception of *L. lactis*). In the absence of *para*-aminobenzoate and folic acid in the medium, steady-state growth was reached in continuous fermentations (Wegkamp, unpublished results), indicating that growth is possible without exogenous *para*-aminobenzoate. We have not yet resolved this apparent gap in the folic acid biosynthesis route.

For thiamine biosynthesis, initially 3 of 10 required reactions were not coupled to a gene. The first gap was related to the phosphorylation of thiamine phosphate (EC 2.7.4.16). The conversion may, however, be catalyzed by the next enzyme in the pathway, thiamine pyrophosphate kinase (lp_1622). The genes encoding these two enzymes, ThiL and THI80, respectively, show significant phylogenetic anticorrelation, which is an indication of analogy (31). The other two gaps in the pathway are related to the desulfuration of cysteine and the incorporation of the liberated sulfur atom into xylulose. Thiamine biosynthesis shares the initial sulfur chemistry with molybdopterin biosynthesis (3, 53). Enteric bacteria use different enzyme sets for the two biosynthesis pathways; i.e., they use different intermediate sulfur carriers (ThiS and MoaD) and different enzymes for adenylating the sulfur carrier (ThiF and MoeB). However, in the *L. plantarum* genome we found only MoaD and MoeB orthologs (Fig. 2A). As the ThiF and MoeB proteins are paralogous and given the similarity in the underlying chemistry (44), we assume that *L. plantarum* MoaD and MoeB are also involved in thiamine biosynthesis.

In the enteric bacteria, the gene products that bind the activated sulfur donor and catalyze the conversion of 1-deoxy-D-xylulose-5-phosphate to 4-methyl-5-(beta-hydroxyethyl)thiazole-phosphate are ThiG and ThiH. In *Bacillus subtilis* and many other bacteria, these proteins are ThiG and ThiO (31, 37, 44). Phylogenetic analysis showed that the *L. plantarum* genome has neither a *thiH* ortholog nor a *thiO* ortholog. The *L. plantarum* gene lp_1776, however, is orthologous to the *B. subtilis yurR* gene, a paralog of the gene encoding ThiO, making lp_1776 a potential candidate for the missing function. Nevertheless, since ThiG and ThiL gene orthologs are still missing from the genome, the actual mechanism of sulfur transfer is still unclear.

Pyridoxal-5-phosphate (vitamin B_6) is a ubiquitous cofactor that is required in many amino acid-converting reactions, including transamination, deamination, and decarboxylation reactions. In the Brenda database (38, 46) 46 enzymes are reported to require pyridoxal-5-phosphate, and 7 of these enzymes are present in *L. plantarum*. Although only salvage enzymes to transform vitamin B_6 precursors (pyridoxine and pyridoxamine) into the active form, pyridoxal-5-phosphate, were identified in the *L. plantarum* genome, pyridoxine and pyridoxamine appeared to be not absolutely required for growth in our omission experiments. Growth was merely inhibited, and pyridoxamine seemed to be the preferred precursor (Table 1). Yet, in a minimal medium with only 11 amino acids (the amino acids indicated by boldface type in Table 1), vitamin B_6 precursors were required for growth (results not shown). Thus, the experimental results indicate that in a complete CDM with all but two amino acids, interconversions of amino acids are not absolutely required for growth and the cells can grow without exogenous pyridoxal-5-phosphate precursors. In minimal medium, however, pyridoxal-5-phosphate-dependent reactions are essential for de novo amino acid biosynthesis, including aspartate biosynthesis (aspartate transaminase [EC 2.6.1.1]), serine biosynthesis (phosphoserine transaminase [EC 2.6.1.52]), and lysine biosynthesis (diaminopimelate decarboxylase [EC 4.1.1.20]).

(ii) Amino acids. We previously reported that most amino acid biosynthesis pathways in *L. plantarum* are complete (26); only the pathways for the biosynthesis of the branched-chain amino acids Ile, Leu, and Val are clearly absent. This was confirmed by the growth experiments; no growth was observed when branched-chain amino acids were omitted. The minor growth on isoleucine observed was due entirely to isoleucine contamination in the other amino acids: The initial isoleucine concentration in the medium (5 mg/liter) was similar to the expected isoleucine content in the final biomass (based on biomass composition experiments) (results not shown).

Although the other amino acid "biosynthesis" routes seemed to be complete, three of the amino acids, arginine, glutamate, and tryptophan, could also not be removed from CDM without serious growth effects (the final optical density was less than 25% of that in complete medium), as shown in Table 1. In a similar study Morishita and colleagues (32) found that seven amino acids are essential for growth of *L. plantarum* ATCC 8014 (the parent strain of *L. plantarum* WCFS1); these amino acids are the same amino acids that we found with the exception of arginine, and methionine and phenylalanine were additional essential amino acids. Omission of the last two amino acids also impaired growth in our study (the final optical den-

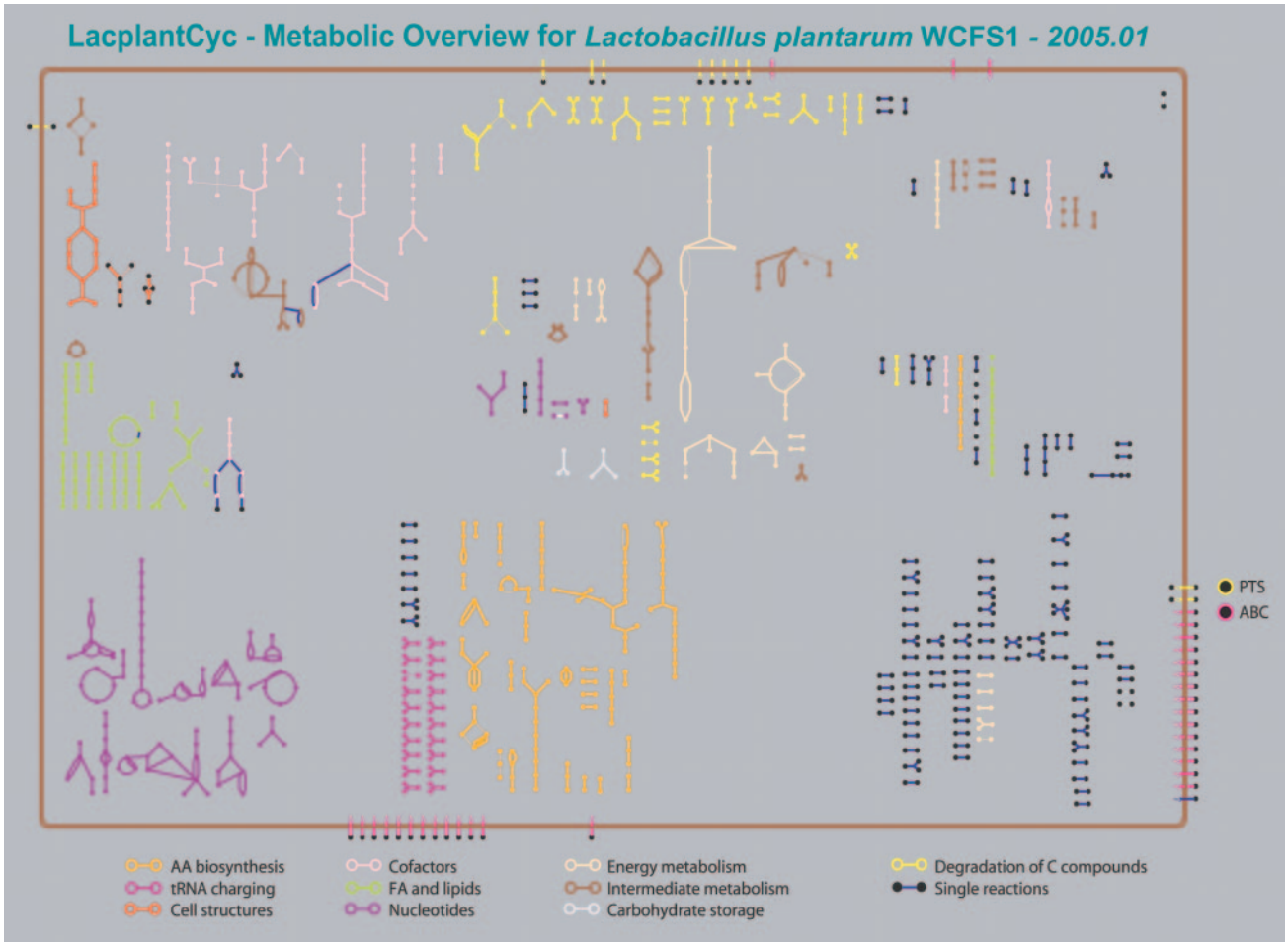

FIG. 4. Metabolic overview map of *L. plantarum*, in which the edges, representing the reactions, are color coded. Different metabolic categories are defined by color for clarity. AA, amino acid; FA, fatty acids.

sity was approximately 40% of that in CDM [Table 1]). The minimal medium that sustained maximal growth in the study of Morishita et al. contained 12 amino acids (32), one more than our minimal medium. Thus, our results agree with those of Morishita et al. very well. The presumed discrepancy between the completeness of the biosynthetic pathway and the observed auxotrophy is discussed below.

The reason that tryptophan is required for growth is most likely the presence of the other aromatic amino acids in the medium. It is known that the biosynthesis of aromatic amino acids is regulated by feedback inhibition by the products. Aromatic amino acids inhibit, among other steps, the 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase (EC 4.1.2.15) reaction, the committed step in chorismate biosynthesis (12). Hence, the presence of phenylalanine and tyrosine in the medium could inhibit the synthesis of tryptophan. The fact that very reasonable growth was observed when tyrosine or phenylalanine was omitted (Table 1) may reflect different relative sensitivities of the enzymes to the different aromatic amino acids; i.e., the biosynthetic enzymes specific for tyrosine and phenylalanine may be less sensitive to feedback inhibition than the enzymes specific for the biosynthesis of tryptophan. This hypothesis remains to be tested experimentally. The observed growth without phenylalanine was quite unexpected, as a required gene in the biosynthetic pathway (encoding prephenate dehydratase [EC 4.2.1.51]) was not found in the genome. We have not been able to close this gap so far.

Arginine is another amino acid for which the pathway seems complete, yet no growth was observed when arginine was omitted from the medium. *Lactobacillus* strains vary in their auxotrophy for arginine; isolates from dairy environments tend to be auxotrophs, whereas isolates from vegetables and humans are mostly prototrophs (5). Since our strain was isolated from human saliva (26), the requirement for arginine in the medium for *L. plantarum* WCFS1 was unexpected, and we do not know whether regulatory effects (feedback mechanisms similar to what was observed for tryptophan) or cumulative mutations in the genes involved in arginine biosynthesis are responsible for the auxotrophy.

Glutamate is synthesized from α-ketoglutarate and NH_3 by glutamate dehydrogenase present in *L. plantarum* (lp_1169), and hence, the glutamate biosynthesis pathway was originally considered "complete." This is an example where cutting a metabolic network into smaller pathways, in this case starting with α-ketoglutarate and ending with glutamate, can be misleading. The auxotrophy for glutamate in LAB is well known

and is caused by an incomplete TCA cycle and hence no supply of α-ketoglutarate. The role of glutamate dehydrogenase is probably related to transamination reactions involved in amino acid degradation and not in glutamate biosynthesis (15, 52). Actually, all reactions in LacplantCyc that could eventually produce glutamate use either α-ketoglutarate or glutamine (which was not present in the medium as a substrate).

The methionine biosynthetic pathway appeared to be complete at first glance, and growth was possible without methionine, although the rate was lower (Table 1). However, upon closer inspection the pathway was annotated to require succinyl coenzyme A as a substrate, and the incomplete TCA cycle does not provide this compound. It is known, however, that most gram-positive bacteria use acetyl-CoA as the substrate (43). The confusion arises from the fact that there is a specific group of genes in the ERGO and KEGG databases that are annotated to encode "homoserine *O*-acetyltransferases," whereas there is another group of genes, including those of *L. plantarum*, *B. subtilis*, and *E. coli*, that are annotated as genes that encode "homoserine *O*-succinyltransferases." Based on phylogenetic analysis, these two groups of genes form distinct orthologous groups (results not shown). Within the group of orthologous genes supposedly encoding "homoserine *O*-succinyltransferases," however, there are genes that have been experimentally shown to have either succinyl-CoA (*E. coli*) or acetyl-CoA (*B. subtilis*) transfer capacity (see references 43 and 45 for overviews and references). The original annotation of *L. plantarum* gene lp_2537 as a homoserine *O*-succinyltransferase gene was therefore changed to annotation as a homoserine *O*-acetyltransferase gene.

Summary of the most important functionalities of LacplantCyc. We offer LacplantCyc to the scientific community as an online reference database (encyclopedia) of metabolic pathways in gram-positive bacteria in general and lactic acid bacteria in particular. This database can be queried via the World Wide Web with reactions, pathways, or genes as inputs. It is also possible to query the database with an uploaded sequence; via an internal BLAST the sequence is compared to the *L. plantarum* genome, and the output contains links to the BioCyc pages for any gene or protein that produces a hit. The corresponding PGDB puts the functional annotation of the isolated genes in a metabolic context. A complete pathway overview can be generated, as shown in Fig. 4. The overview allows several comparisons by color coding of the edges. Once the connections between genes, proteins, and reactions have been defined, any numerical attribute of reactions can be projected on the metabolic map. The attribute could be, for example, the number of elementary modes going through an enzyme (49), the metabolic capabilities of the organism compared to those of another organism, or microarray data.

In summary, LacplantCyc provides researchers with a helpful tool for analysis of the functional genomic information for *L. plantarum*. The interactive metabolic map allows visualization of the metabolic context and, as such, complements tools for visualization of genomic context, such as the recently developed Microbial Genome Viewer (22). A link to Microbial Genome Viewer is provided in LacplantCyc. Visualization of data sets at different levels of detail is essential to help interpret the data from a biological point of view, and we hope that LacplantCyc will be a useful tool for doing this.

ACKNOWLEDGMENTS

We thank Jos Boekhorst, Richard Notebaart, and Robert Kerkhoven for their help with the pathway reconstruction and figures. We also thank the team of Peter Karp for their support concerning Pathway Tools issues.

This work took place in the Kluyver Centre for Genomics of Industrial Fermentations and was supported by grant NWO-BMI 050.50.206.

REFERENCES

1. **Altermann, E., W. M. Russell, M. A. Azcarate-Peril, R. Barrangou, B. L. Buck, O. McAuliffe, N. Souther, A. Dobson, T. Duong, M. Callanan, S. Lick, A. Hamrick, R. Cano, and T. R. Klaenhammer.** 2005. Complete genome sequence of the probiotic lactic acid bacterium *Lactobacillus acidophilus* NCFM. Proc. Natl. Acad. Sci. USA **102:**3906–3912.
2. **Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock.** 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. **25:**25–29.
3. **Begley, T. P., D. M. Downs, S. E. Ealick, F. W. McLafferty, A. P. Van Loon, S. Taylor, N. Campobasso, H. J. Chiu, C. Kinsland, J. J. Reddick, and J. Xi.** 1999. Thiamine biosynthesis in prokaryotes. Arch. Microbiol. **171:**293–300.
4. **Boels, I. C., R. Van Kranenburg, M. W. Kanning, B. F. Chong, W. M. De Vos, and M. Kleerebezem.** 2003. Increased exopolysaccharide production in *Lactococcus lactis* due to increased levels of expression of the NIZO B40 *eps* gene cluster. Appl. Environ Microbiol. **69:**5029–5031.
5. **Bringel, F., and J. C. Hubert.** 2003. Extent of genetic lesions of the arginine and pyrimidine biosynthetic pathways in *Lactobacillus plantarum*, *L. paraplantarum*, *L. pentosus*, and *L. casei*: prevalence of CO_2-dependent auxotrophs and characterization of deficient *arg* genes in *L. plantarum*. Appl. Environ. Microbiol. **69:**2674–2683.
6. **Brown, G. M., and J. M. Williamson.** 1982. Biosynthesis of riboflavin, folic acid, thiamine, and pantothenic acid. Adv. Enzymol. Relat. Areas Mol. Biol. **53:**345–381.
7. **Burgess, C., M. O'Connell-Motherway, W. Sybesma, J. Hugenholtz, and D. van Sinderen.** 2004. Riboflavin production in *Lactococcus lactis*: potential for in situ production of vitamin-enriched foods. Appl. Environ. Microbiol. **70:**5769–5777.
8. **Covert, M. W., E. M. Knight, J. L. Reed, M. J. Herrgard, and B. O. Palsson.** 2004. Integrating high-throughput and computational data elucidates bacterial networks. Nature **429:**92–96.
9. **Davidson, A. L., and J. Chen.** 2004. ATP-binding cassette transporters in bacteria. Annu. Rev. Biochem. **73:**241–268.
10. **de Vos, W. M., and J. Hugenholtz.** 2004. Engineering metabolic highways in lactococci and other lactic acid bacteria. Trends Biotechnol. **22:**72–79.
11. **Edgar, R. C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32:**1792–1797.
12. **Euverink, G. J., G. I. Hessels, C. Franke, and L. Dijkhuizen.** 1995. Chorismate mutase and 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase of the methylotrophic actinomycete *Amycolatopsis methanolica*. Appl. Environ. Microbiol. **61:**3796–3803.
13. **Green, M. L., and P. D. Karp.** 2004. A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. BMC Bioinformatics **5:**76.
14. **Gutknecht, R., R. Beutler, L. F. Garcia-Alles, U. Baumann, and B. Erni.** 2001. The dihydroxyacetone kinase of *Escherichia coli* utilizes a phosphoprotein instead of ATP as phosphoryl donor. EMBO J. **20:**2480–2486.
15. **Helinck, S., D. Le Bars, D. Moreau, and M. Yvon.** 2004. Ability of thermophilic lactic acid bacteria to produce aroma compounds from amino acids. Appl. Environ. Microbiol. **70:**3855–3861.
16. **Hucka, M., A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, Goryanin II, W. J. Hedley, T. C. Hodgman, J. H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novere, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang.** 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics **19:**524–531.
17. **Hugenholtz, J., M. Kleerebezem, M. Starrenburg, J. Delcour, W. de Vos, and P. Hols.** 2000. *Lactococcus lactis* as a cell factory for high-level diacetyl production. Appl. Environ. Microbiol. **66:**4112–4114.
18. **Hugenholtz, J., W. Sybesma, M. N. Groot, W. Wisselink, V. Ladero, K. Burgess, D. van Sinderen, J. C. Piard, G. Eggink, E. J. Smid, G. Savoy, F. Sesma, T. Jansen, P. Hols, and M. Kleerebezem.** 2002. Metabolic engineering of lactic acid bacteria for the production of nutraceuticals. Antonie Leeuwenhoek **82:**217–235.

19. **Karp, P. D., S. Paley, and P. Romero.** 2002. The Pathway Tools software. Bioinformatics **18**(Suppl. 1)**:**S225–S232.
20. **Karp, P. D., M. Riley, S. M. Paley, and A. Pellegrini-Toole.** 2002. The MetaCyc Database. Nucleic Acids Res. **30:**59–61.
21. **Karp, P. D., M. Riley, M. Saier, I. T. Paulsen, J. Collado-Vides, S. M. Paley, A. Pellegrini-Toole, C. Bonavides, and S. Gama-Castro.** 2002. The EcoCyc Database. Nucleic Acids Res. **30:**56–58.
22. **Kerkhoven, R., F. H. van Enckevort, J. Boekhorst, D. Molenaar, and R. J. Siezen.** 2004. Visualization for genomics: the Microbial Genome Viewer. Bioinformatics **20:**1812–1814.
23. **Keseler, I. M., J. Collado-Vides, S. Gama-Castro, J. Ingraham, S. Paley, I. T. Paulsen, M. Peralta-Gil, and P. D. Karp.** 2005. EcoCyc: a comprehensive database resource for *Escherichia coli*. Nucleic Acids Res. **33:**D334–337.
24. **Klaenhammer, T., E. Altermann, F. Arigoni, A. Bolotin, F. Breidt, J. Broadbent, R. Cano, S. Chaillou, J. Deutscher, M. Gasson, M. van de Guchte, J. Guzzo, A. Hartke, T. Hawkins, P. Hols, R. Hutkins, M. Kleerebezem, J. Kok, O. Kuipers, M. Lubbers, E. Maguin, L. McKay, D. Mills, A. Nauta, R. Overbeek, H. Pel, D. Pridmore, M. Saier, D. van Sinderen, A. Sorokin, J. Steele, D. O'Sullivan, W. de Vos, B. Weimer, M. Zagorec, and R. Siezen.** 2002. Discovering lactic acid bacteria by genomics. Antonie Leeuwenhoek **82:**29–58.
25. **Klaus, S. M., A. Wegkamp, W. Sybesma, J. Hugenholtz, J. F. Gregory III, and A. D. Hanson.** 2005. A nudix enzyme removes pyrophosphate from dihydroneopterin triphosphate in the folate synthesis pathway of bacteria and plants. J. Biol. Chem. **280:**5274–5280.
26. **Kleerebezem, M., J. Boekhorst, R. van Kranenburg, D. Molenaar, O. P. Kuipers, R. Leer, R. Tarchini, S. A. Peters, H. M. Sandbrink, M. W. Fiers, W. Stiekema, R. M. Lankhorst, P. A. Bron, S. M. Hoffer, M. N. Groot, R. Kerkhoven, M. de Vries, B. Ursing, W. M. de Vos, and R. J. Siezen.** 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1. Proc. Natl. Acad. Sci. USA **100:**1990–1995.
27. **Kleerebezem, M., I. C. Boels, M. N. Groot, I. Mierau, W. Sybesma, and J. Hugenholtz.** 2002. Metabolic engineering of *Lactococcus lactis*: the impact of genomics and metabolic modelling. J. Biotechnol. **98:**199–213.
28. **Kleerebezem, M., P. Hols, and J. Hugenholtz.** 2000. Lactic acid bacteria as a cell factory: rerouting of carbon metabolism in *Lactococcus lactis* by metabolic engineering. Enzyme Microb. Technol **26:**840–848.
29. **Krieger, C. J., P. Zhang, L. A. Mueller, A. Wang, S. Paley, M. Arnaud, J. Pick, S. Y. Rhee, and P. D. Karp.** 2004. MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res. **32:**D438–442.
30. **Molin, G.** 2001. Probiotics in foods not containing milk or milk constituents, with special reference to *Lactobacillus plantarum* 299v. Am. J. Clin. Nutr. **73:**380S–385S.
31. **Morett, E., J. O. Korbel, E. Rajan, G. Saab-Rincon, L. Olvera, M. Olvera, S. Schmidt, B. Snel, and P. Bork.** 2003. Systematic discovery of analogous enzymes in thiamine biosynthesis. Nat. Biotechnol. **21:**790–795.
32. **Morishita, T., Y. Deguchi, M. Yajima, T. Sakurai, and T. Yura.** 1981. Multiple nutritional requirements of lactobacilli: genetic lesions affecting amino acid biosynthetic pathways. J. Bacteriol. **148:**64–71.
33. **Oku, Y., K. Kurokawa, N. Ichihashi, and K. Sekimizu.** 2004. Characterization of the *Staphylococcus aureus mprF* gene, involved in lysinylation of phosphatidylglycerol. Microbiology **150:**45–51.
34. **O'Leary, W. M., and S. G. Wilkinson.** 1988. Gram-positive bacteria, p. 117–195. *In* C. Ratledge and S. G. Wilkinson (ed.), Microbial lipids, vol. 1. Academic Press, London, United Kingdom.
35. **Osterman, A., and R. Overbeek.** 2003. Missing genes in metabolic pathways: a comparative genomics approach. Curr. Opin. Chem. Biol. **7:**238–251.
36. **Palsson, B. O.** 2004. In silico biotechnology. Era of reconstruction and interrogation. Curr. Opin. Biotechnol. **15:**50–51.
37. **Park, J. H., P. C. Dorrestein, H. Zhai, C. Kinsland, F. W. McLafferty, and T. P. Begley.** 2003. Biosynthesis of the thiazole moiety of thiamine pyrophosphate (vitamin B11). Biochemistry **42:**12430–12438.
38. **Pharkya, P., E. V. Nikolaev, and C. D. Maranas.** 2003. Review of the BRENDA Database. Metab. Eng. **5:**71–73.
39. **Poolman, B., and W. N. Konings.** 1988. Relation of growth of *Streptococcus lactis* and *Streptococcus cremoris* to amino acid transport. J. Bacteriol. **170:** 700–707.
40. **Postma, P. W., J. W. Lengeler, and G. R. Jacobson.** 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. Microbiol. Rev. **57:**543–594.
41. **Pridmore, R. D., B. Berger, F. Desiere, D. Vilanova, C. Barretto, A. C. Pittet, M. C. Zwahlen, M. Rouvet, E. Altermann, R. Barrangou, B. Mollet, A. Mercenier, T. Klaenhammer, F. Arigoni, and M. A. Schell.** 2004. The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. Proc. Natl. Acad. Sci. USA **101:**2512–2517.
42. **Rizzi, M., and H. Schindelin.** 2002. Structural biology of enzymes involved in NAD and molybdenum cofactor biosynthesis. Curr. Opin. Struct. Biol. **12:** 709–720.
43. **Rodionov, D. A., A. G. Vitreschak, A. A. Mironov, and M. S. Gelfand.** 2004. Comparative genomics of the methionine metabolism in Gram-positive bacteria: a variety of regulatory systems. Nucleic Acids Res. **32:**3340–3353.
44. **Rodionov, D. A., A. G. Vitreschak, A. A. Mironov, and M. S. Gelfand.** 2002. Comparative genomics of thiamine biosynthesis in procaryotes. New genes and regulatory mechanisms. J. Biol. Chem. **277:**48949–48959.
45. **Saint-Girons, I., C. Parsot, M. M. Zakin, O. Barzu, and G. N. Cohen.** 1988. Methionine biosynthesis in *Enterobacteriaceae*: biochemical, regulatory, and evolutionary aspects. Crit. Rev. Biochem. **23**(Suppl. 1)**:**S1–S42.
46. **Schomburg, I., A. Chang, C. Ebeling, M. Gremse, C. Heldt, G. Huhn, and D. Schomburg.** 2004. BRENDA, the enzyme database: updates and major new developments. Nucleic Acids Res. **32:**D431–433.
47. **Siezen, R. J., F. H. van Enckevort, M. Kleerebezem, and B. Teusink.** 2004. Genome data mining of lactic acid bacteria: the impact of bioinformatics. Curr. Opin. Biotechnol. **15:**105–115.
48. **Steidler, L.** 2001. *Lactococcus lactis*, a tool for the delivery of therapeutic proteins treatment of IBD. Sci. World J. **1:**216–217.
49. **Stelling, J., S. Klamt, K. Bettenbrock, S. Schuster, and E. D. Gilles.** 2002. Metabolic network structure determines key aspects of functionality and regulation. Nature **420:**190–193.
50. **Suzuki, Y., and G. M. Brown.** 1974. The biosynthesis of folic acid. XII. Purification and properties of dihydroneopterin triphosphate pyrophosphohydrolase. J. Biol. Chem. **249:**2405–2410.
51. **Sybesma, W., M. Starrenburg, M. Kleerebezem, I. Mierau, W. M. de Vos, and J. Hugenholtz.** 2003. Increased production of folate by metabolic engineering of *Lactococcus lactis*. Appl. Environ. Microbiol. **69:**3069–3076.
52. **Tanous, C., A. Kieronczyk, S. Helinck, E. Chambellon, and M. Yvon.** 2002. Glutamate dehydrogenase activity: a major criterion for the selection of flavour-producing lactic acid bacteria strains. Antonie Leeuwenhoek **82:**271–278.
53. **Taylor, S. V., N. L. Kelleher, C. Kinsland, H. J. Chiu, C. A. Costello, A. D. Backstrom, F. W. McLafferty, and T. P. Begley.** 1998. Thiamine biosynthesis in *Escherichia coli*. Identification of this thiocarboxylate as the immediate sulfur donor in the thiazole formation. J. Biol. Chem. **273:**16555–16560.
54. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.
55. **Vergnes, A., K. Gouffi-Belhabich, F. Blasco, G. Giordano, and A. Magalon.** 2004. Involvement of the molybdenum cofactor biosynthetic machinery in the maturation of the *Escherichia coli* nitrate reductase A. J. Biol. Chem. **279:**41398–41403.